

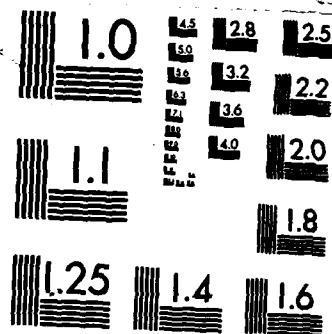
ADJUSTED BELIEF STRUCTURES(U) WISCONSIN UNIV-MADISON
MATHEMATICS RESEARCH CENTER M GOLDSTEIN MAR 85
MRC-TSR-2804 DAAG29-80-C-0041

UNCLASSIFIED

F/G 5/10

NL

[illegible]



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

7

AD-A154 804

MRC Technical Summary Report #2804

ADJUSTED BELIEF STRUCTURES

Michael Goldstein

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

March 1985

(Received August 9, 1984)

DTIC FILE COPY

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

DTIC
ELECTE
JUN 11 1985
S G D

85 06 10 172

Accession For	
DTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A/1	

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

ADJUSTED BELIEF STRUCTURES

Michael Goldstein*

Technical Summary Report #2804

March 1985

ABSTRACT



→ In the study of the foundations of the subjectivist theory of statistics, we find that each aspect of the theory corresponds to a different application of a single manipulation, namely the adjustment of one belief structure by another belief structure. This article describes the technical machinery of this manipulation to a sufficient level of detail to cover all of the applications of the adjusted belief structure in the foundations of the theory. We also discuss the relationship between the adjustment of belief structures and the conditioning of random variables. (Essentially the latter is a simple special case of the former.)

AMS (MOS) Subject Classifications: Primary 62A15; Secondary 60A99

→ Key Words: Coherence, conditional prevision, prevision, projection. ↗

Work Unit Number 4 - Statistics and Probability

*

University of Hull, Department of Statistics, Hull, England HU6 7RX.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

Adjusted Belief Structures

Michael Goldstein

1. Introduction
2. Subspaces and alternative inner products
3. Adjusted belief structures
4. Adjusted belief structures and projections
5. Conditional prevision
6. Adjusted belief structures and conditional beliefs
7. Adjusted belief construction

ADJUSTED BELIEF STRUCTURES

Michael Goldstein*

1. Introduction

Probability theory has an extremely wide range of application. However, in most applications, the basic manipulations are essentially the same, the differences arising from the context within which these manipulations are expressed. Thus, identifying and understanding these basic manipulations is essential for separating out the underlying conceptual issues from the specific technical difficulties of a particular application.

This is particularly important in the study of the foundations, as our aim is to describe the possibilities of the theory. Thus, we shall now identify, and discuss in detail, the single manipulation of the belief structure which we repeatedly require, namely the adjustment of one belief structure by another belief structure. Just as we find that our study of the foundations is naturally expressed in terms of belief structures, so we will find in subsequent articles that each aspect of the theory will simply correspond to a different adjustment of the belief structure.

This article is a sequel to the previous technical report entitled "Belief Structures", and is the second of a series of articles laying the foundation for the subjectivist theory. Our intention in this article is simply to explain the general process of adjustment in sufficient detail to cover all of the various applications of this process that we make in subsequent articles. The notation is as in the previous report.

In order to motivate the construction, we will begin by discussing the simplest example of such an adjustment.

*
University of Hull, Department of Statistics, Hull, England HU6 7RX.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

2. Subspaces and alternative inner products

In our investigations, we choose as fundamental the inner product space A defined by the inner product $(X,Y) = P(XY)$ over the linear space L . DeFinetti (Theory of Probability, 1974, section 4.17) discusses this space as a simple geometric interpretation of the prevision of products of random variables. He then observes that there is an alternative inner product which is more commonly used, and may appear more natural. This inner product $(\cdot, \cdot)'$ is defined over L by

$$(X,Y)' = P(X-P(X))(Y-P(Y)) = \text{cov}(X,Y) .$$

Let us call the inner product space generated by this inner product A' . (Note that we must identify, as equivalence classes, all random quantities which differ by a constant, so that, for example, X_0 (the unit constant) is equivalent to the zero vector in A' .)

Thus, in A' , $|X|$ is the standard deviation of X . Very loosely, we can consider that vectors with large norms in A' correspond to random quantities whose values you are "very uncertain of", while vectors with small norm are those whose value you are "fairly sure of". The inner product is covariance, and orthogonality corresponds to zero correlation.

Thus, if we wish to consider "relationships" between random quantities, and to express our "degree of uncertainty" about these quantities, then in many ways the space A' seems a more natural object than the space A . We will show that, in a certain sense, A' is not a "different" inner product space to A , but that A' can more usefully be considered as an "adjustment" of A . This is a typical example of the purpose of adjustment, namely to remove certain features of the space A which are not of immediate interest (such as the individual previsions of the elements of C), in order to focus attention on aspects which are of interest (such as the "uncertainties", or variances, of the elements of C).

As a first step in describing our construction, notice that rather than defining A and A' in terms of different inner products, we can instead view A' as a subspace of A . Specifically, define A^* to be the orthogonal complement of the subspace A_0 of A , (A_0 is the subspace spanned by X_0). Equivalently A^* is the subspace of A of all vectors of the form $X - P(X)X_0$. We can now identify the spaces A' and A^* , simply

by identifying each (equivalence class) X with the corresponding random quantity "corrected for the mean", i.e. $X - P(X)X_0$. (Note that as members of an equivalence class differ by a constant, each member of a particular class is identified with the same random quantity.) The identification $T(X) = X - P(X)X_0$ preserves the inner product, as for any $X, Y \in A$,

$$(X, Y)' = (T(X), T(Y)) .$$

Thus, we will never require the alternative inner product structure A' as whenever we want only to consider variation about the mean, then we can focus attention on the subspace A^* of A . Notice in particular that A is the orthogonal sum of A^* and A_0 (i.e. $A = A_0 \oplus A^*$).

3. Adjusted belief structures

The construction of section 2 is useful in its own right, but it is also the simplest example of a very general construction. In this example, we began with a belief structure constructed around a list $C = \{X_1, \dots, X_m\}$. We then introduced a new quantity X_0 into the structure and used this new quantity to split the belief structure into orthogonal subspaces. The general construction, of which this is a special case is as follows.

- (i) We begin with a belief structure A , constructed from a collection $C = \{X_0, X_1, X_2, \dots, X_m\}$.
- (ii) We introduce a further collection of random quantities $C' = \{Y_1, \dots, Y_k\}$ (where some elements of C and C' may be the same).
- (iii) We construct the belief structure B from the collection C' (i.e. we evaluate $(Y_i, Y_j) = P(Y_i Y_j)$ for each i, j).
- (iv) We now add the belief structures A and B , to give a new belief structure $D = A + B$, spanned by the elements $\{X_0, \dots, X_m, Y_1, \dots, Y_k\}$ (i.e. we evaluate each $P(X_i Y_j)$).
- (v) We now divide the space D into two orthogonal subspaces B and B^\perp , where B^\perp is the orthogonal complement of B in D (i.e. so that $D = B \oplus B^\perp$).

In the construction of section 2, the collection C' was the single quantity X_0 . The belief structure B was the space we previously termed A_0 . When we constructed the space A by adding A_0 to the space spanned by $\{X_1, \dots, X_m\}$, we evaluated $(X_0, X_1) = P(X_1)$ for each i . We then divided A into orthogonal subspaces A_0 and A_0^\perp , where A_0^\perp is the space we previously termed A^* .

In this case, introducing the constant X_0 as a subspace has separated out your beliefs so that you may, if you wish, consider variances and covariances separately from mean values. The general construction will be useful whenever the new spaces B and B^\perp that are created have a natural subjective interpretation. The full importance of this construction will be revealed when we consider, in subsequent articles, the revision of your beliefs. We will show that, in a certain important sense, the separation between subspaces is preserved under the revision of the inner product over the belief structure, for a wide class of choices of B . Further, we will identify a particular choice of space B for which the above construction will essentially define the properties of your revisions of belief.

However, for the present, let us simply note that this is an interesting construction, which we are likely to use fairly often. Thus, we introduce a helpful piece of notation to describe the construction.

Definition. If A and B are both belief structures, then the belief structure A adjusted for the belief structure B , written A/B , is defined to be the orthogonal complement B^\perp of the subspace B in the space $A + B$.

Thus, for example, $A^* = A/A_0$, i.e. A^* is the space A adjusted for the constant space. Notice that the definition is the same whether the elements of A and B are partially or completely distinct. The notation identifies the orthogonal complement of a space with the associated quotient space. It is suggestive of a connection between the operations of adjusting a space and the "conditioning" of random quantities. This connection will be explored after we have briefly outlined some of the properties of adjusted spaces.

4. Adjusted belief structures and projections

Adjusted belief structures obey a few simple rules which we list here for convenience. (The proofs are straightforward.) For any belief structures A_1, \dots, A_k, B ,

- (i) $A_1/B = 0$, the zero space, if and only if $A_1 \subseteq B$.
- (ii) $A_1/B = A_1$ if and only if $A_1 \perp B$
- (iii) $(A_1 + A_2)/B = (A_1/B) + (A_2/B)$
- (iv) $A_1 + A_2 + \dots + A_k = A_1 \oplus (A_2/A_1) \oplus (A_3/(A_1 + A_2)) \oplus \dots \oplus (A_k/A_1 + \dots + A_{k-1})$
- (v) $(A_1/B) \perp (A_2/B)$ if and only if $A_1 \subseteq B + D_1, A_2 \subseteq B + D_2$, where B, D_1, D_2 are mutually orthogonal.

(Property (iv) is useful when we wish to systematically adjust each of a collection of belief structures. Property (v) is the key to the general representation theorems that we shall develop in later articles.)

The general properties of adjusted structures are, however, more conveniently expressed by linking each subspace with the corresponding orthogonal projection into the subspace.

Notation. For any closed space B , we denote by $P_B(\cdot)$, the orthogonal projection operator from A into B (i.e., for each $X \in A$, $P_B(X)$ is the choice of element $Y \in B$, for which $\|X - Y\|$ is minimized over all $Y \in B$).

We do not require that B should be a subspace of A . Thus, the first stage in constructing $P_B(\cdot)$ is to construct the combined space $D = A + B$. The orthogonal projection operator into B is defined over D , and P_B is the restriction of this operator to A (now considered as a subspace of D). Notice in particular that P_B is the identity operator if and only if $A \subseteq B$ and P_B is the zero operator if and only if $A \perp B$.

The relationship between projections and adjusted beliefs is that, for any spaces B_1 and B_2 , we have

$$P_{(B_1+B_2)} = P_{B_1} + P_{(B_2/B_1)} \quad (1)$$

Thus, we can add a further basic property to the properties (i) - (v) of adjusted spaces listed above, namely

(vi) For any belief structures A, B_1, B_2 ,

$$A/(B_1+B_2) = (A/B_1)/(B_2/B_1) .$$

(The space $A/(B_1+B_2)$ is spanned by elements of the form $X - P_{B_1+B_2}X = X - P_{B_1}X - P_{(B_2/B_1)}X$. The space A/B_1 is spanned by elements of the form $X - P_{B_1}X$, so that $(A/B_1)/(B_2/B_1)$ is spanned by elements of the form $X - P_{B_1}X - P_{B_2/B_1}X + P_{B_2/B_1}P_{B_1}X$, which is the same as the elements of $A/(B_1+B_2)$ as $P_{B_2/B_1}P_{B_1}$ is the zero operator (because $B_1, B_2/B_1$ are orthogonal). The simplest special case of (vi) is when $B_1 \perp B_2$, so that

$$(vii) A/B_1 \oplus B_2 = (A/B_1)/B_2 = (A/B_2)/B_1 .$$

It will often be useful to be able to "adjust" spaces in several stages, and so this raises a natural converse question to property (vii) namely for what spaces B_1, B_2 does

$$(A/B_1)/B_2 = (A/B_2)/B_1 ,$$

and when does either adjustment correspond to a single adjustment (A/D) for some further space D ? The answer is as follow

$$(viii) (A/B_1)/B_2 = (A/B_2)/B_1$$

if and only if P_{B_1} and P_{B_2} are commuting projections (i.e. $P_{B_1}P_{B_2} = P_{B_2}P_{B_1}$).

P_{B_1} and P_{B_2} commute if $B_1 = B \oplus D_1, B_2 = B \oplus D_2$, where B, D_1, D_2 are mutually orthogonal. (This condition is trivially satisfied when $B_1 \perp B_2$.) In this case $P_{B_1}P_{B_2} = P_{B_2}P_{B_1} = P_B$ and $(A/B_1)/B_2 = (A/B_2)/B_1 = A/(B \oplus D_1 \oplus D_2)$.

Further, as B_1 and (B_2/B_1) are orthogonal spaces, we can automatically decompose the inner product over $B_1 + B_2$. For any $X, Y \in A$, we have

$$(P_{(B_1+B_2)}(X), P_{(B_1+B_2)}(Y)) = (P_{B_1}(X), P_{B_1}(Y)) + (P_{(B_2/B_1)}(X), P_{(B_2/B_1)}(Y)) . \quad (2)$$

A special case of this decomposition which we will frequently use follows from setting $B_2 = A$. Thus, $P_{(B+A)}$ is the identity operator and so each choice of space B resolves each element $X \in A$ into two orthogonal components, as

$$X = P_B(X) + P_{A/B}(X) \quad (3)$$

so that the inner product over A is decomposed as

$$(X, Y) = (P_B(X), P_B(Y)) + (P_{A/B}(X), P_{A/B}(Y)) . \quad (4)$$

Thus, when you construct the inner product structure over A , you may separately consider your inner product structure B and your belief structure (A/B) and these assessments must combine to determine your belief structure A . This will provide a wide variety of coherence checks for your assessments. (We will see in later articles that the most important coherence checks will be associated with your revisions of belief over B and over (A/B) .) These checks will be relevant when either B or (A/B) or both have a natural interpretation. We have seen one example already, namely in section 2. Here $A^* = A/A_0$, and

$$P_{A_0}(X) = P(X)X_0, \quad P_{A^*}(X) = X - P(X)X_0 ,$$

so that relation (1) becomes

$$(X, Y) = P(X)P(Y) + \text{cov}(X, Y) ,$$

i.e. we have decomposed the inner product into a separate consideration of means and covariances.

In the previous report on "Belief Structures" we noted the fundamental relationship between prevision and projection. Notice that we can numerically identify the prevision of X , $P(X)$, with the particular projection $P_{A_0}(X)$. Further, the projection P_{A_0} is simply a particular case of the general projection. P_B is, in a sense, a generalization of conditional prevision, with the space B acting analogously to the "conditioning" random quantities. (We will make the relationship precise in the next subsection.) Just as our choice of notation $P(\cdot)$ allows us to pass interchangeably between probabilities and expectations, so it also allows us to pass interchangeably between previsions and projections.

The space A/A_0 describes the variances and covariances of the elements of A , i.e. the variation in the quantities around the plane of certainty. In the same way A/B summarizes the variation in the elements of A when we have taken account of the variation in B . In a sense $P_{A/B}$ gives the "residual vectors" for the "fitted regression" P_B .

For completeness, we now record the basic formulae relating to projections. We introduce the following piece of notation. Suppose that $a_1, \dots, a_k, b_1, \dots, b_r$ are elements of the inner product space \mathcal{V} . Suppose we write $\underline{a} = (a_1, \dots, a_k), \underline{b} = (b_1, \dots, b_r)$. Then we shall denote by $(\underline{a}|\underline{b})$ the $k \times r$ matrix whose $(i,j)^{\text{th}}$ term is (a_i, b_j) . With this notation, for any finite dimensional subspace \mathcal{B} and any basis of \mathcal{B} , b_1, \dots, b_k , the projection operator $P_{\mathcal{B}}$ can be written, for each X , as

$$P_{\mathcal{B}}(X) = \underline{b}(\underline{b}|\underline{b})^{-1}(\underline{b}|X)$$

where $\underline{b} = (b_1, \dots, b_k)$.

Further, the squared distance between X and $P_{\mathcal{B}}(X)$ is given by the ratio of two determinants, as

$$\|P_{\mathcal{B}}(X) - X\|^2 = \|X - P_{\mathcal{B}}(X)\|^2 = \frac{|(\underline{b}(X)|\underline{b}(X))|}{|(\underline{b}|\underline{b})|}$$

where $\underline{b}(X)$ is the vector (X, b_1, \dots, b_k) .

If b_1, \dots, b_k are an orthogonal basis, (i.e. $(b_i, b_j) = 0, i \neq j$), then the above formulae simplify to give

$$\begin{aligned} P_{\mathcal{B}}(X) &= \sum_i \frac{(b_i, X)}{(b_i, b_i)} b_i \\ \|P_{\mathcal{B}}(X) - X\|^2 &= (X, X) - \sum_i \frac{(b_i, X)^2}{(b_i, b_i)} \end{aligned} \quad (5)$$

Finally, the following structural properties of projection operators will be important in later developments.

(i) Projections are idempotent (i.e. $P_{\mathcal{B}}^2 = P_{\mathcal{B}}$)

(more generally, $P_{\mathcal{B}}P_{\mathcal{B}_1} = P_{\mathcal{B}}$ if $\mathcal{B} \subset \mathcal{B}_1$).

(ii) Projections are self-adjoint, i.e. for any $X, Y \in \mathcal{A} + \mathcal{B}$

$$(X, P_{\mathcal{B}}Y) = (P_{\mathcal{B}}X, Y) .$$

(Note an operator is a projection if and only if it is idempotent and self-adjoint.)

An important consequence of (i) and (ii) is that for any $X, Y \in \mathcal{A} + \mathcal{B}$,

$$\begin{aligned}
(iii) \quad (X, P_B Y) &= (X, P_B P_B Y) \\
&= (P_B X, P_B Y) \\
&= (P_B X, Y) ,
\end{aligned}$$

which, for example, gives a direct demonstration of the relationship

$$\|P_{A/B}(X)\|^2 = \|X\|^2 - \|P_B X\|^2 .$$

(iv) P_B is a bounded linear operator, $\|P_B\| = 1$ over $(A+B)$, so that the restriction of P_B to A has norm not greater than one. (For any linear transformation T , $\|T\| = \sup_{X \neq 0} \frac{\|TX\|}{\|X\|}$.)

5. Conditional prevision

We will now discuss the formal relationship between conditional beliefs and adjusted belief spaces. Thus, we begin by briefly reviewing the notion of conditional probability, or in the present case conditional prevision which De Finetti defines as follows.

Definition. The conditional prevision of the random quantity X , given event H , written $P(X|H)$, is the value that you would choose if, having made this choice, you were to suffer a penalty L given

$$L = K H(X - \bar{x})^2 ,$$

where K defines the units of loss and H is the indicator function for the event H . (In other words, we have a "called-off" penalty, which is only invoked if H occurs.)

The point to observe about the definition is that it appears to make events "special" again. That is, having argued, in constructing the belief structure in our previous report, why we do not need to distinguish at a fundamental level between probability and expectation, we have now introduced a definition which only makes sense when the conditioning random quantity H is a two valued random quantity. If this is really the case, and if our new definition is actually necessary to our subsequent development, then this suggests that such a distinction is indeed crucial to the theory. Further it suggests that the problems that we intended to overcome by working with expectation rather than probability are actually unavoidable, as even if we can avoid constructing an exhaustive collection of outcomes for the primary quantities of interest, we will still be forced to

reduce observational evidence into a partition (which is an even more daunting prospect, as at least you are free to choose your primary quantities of interest, but the "data" is far less under your control). What we will argue in detail in subsequent articles is that events are not "special", and that the restriction of the definition of conditional prevision to events is precisely as arbitrary as would be a restriction of the definition of prevision itself to events.

The coherence condition that De Finetti imposes is that you do not prefer a given penalty if you can choose a different penalty which is certainly smaller. De Finetti shows that the necessary and sufficient condition for coherence in evaluating $P(X|H)$, $P(XH)$ and $P(H)$ is that

$$P(HX) = P(X|H)P(H) ,$$

in addition to the inequality $\inf(X|H) \leq P(X|H) \leq \sup(X|H)$, where the \inf and \sup are over all values of X consistent with H . (Notice that if X is itself an event, then the above condition is the usual theorem of compound probabilities.)

Observe in particular that $P(X|H)$, by definition, expresses your choice made now, before H has been revealed, when confronted with a penalty in X which will be called off unless H occurs. It is very easy to twist this around and declare that if you discover that H has occurred, then your prevision for X "should" become the value that you have assigned for $P(X|H)$. Indeed, all of Bayesian statistics is based around this "principle". We will discuss in detail in a later article, precisely why this view is misguided. For now, let us simply observe that the called off penalty definition of conditional prevision does not, of itself, say anything concerning your future beliefs. Thus, any linkage between conditional prevision and future beliefs is not self evident, but requires additional justification.

Finally, let us briefly outline a useful property of conditioning. Consider any finite partition of possibilities, i.e. a set E_1, \dots, E_k of events such that one and only one of the events will occur. We may define the prevision of X conditional on the partition $\Pi = \{E_1, \dots, E_k\}$, as

$$P(X|\Pi) = P(X|E_1)E_1 + \dots + P(X|E_k)E_k .$$

(In other words, $P(X|\Pi)$ is the random quantity which takes value $P(X|E_i)$ if E_i occurs.) Thus

$$\begin{aligned} P(P(X|\Pi)) &= P(X|E_1)P(E_1) + \dots + P(X|E_k)P(E_k) \\ &= P(XE_1) + \dots + P(XE_k) \\ &= P(X(E_1 + \dots + E_k)) = P(X) \\ &\text{as } E_1 + \dots + E_k = 1. \end{aligned}$$

If Y is the random quantity which takes a finite number of possible values y_1, \dots, y_k , then we can similarly write $P(X|Y) = P(X|\Pi)$ where the partition is over $\{E_i : Y = y_i\}$, and again

$$P(P(X|Y)) = P(X).$$

Does this relationship hold, if we allow Y to take an infinite number of possible values? Yes, if X is bounded and $\sum_{i=1}^{\infty} P(E_i) = 1$, (as we can define $F_n = E_1 + \dots + E_n$, $G_n = 1 - F_n$, and write

$$P(X) = P(X|F_n)P(F_n) + P(X|G_n)P(G_n).$$

The second term on the right hand side tends to zero, while the first tends to $P(P(X|Y))$.

However, in general, when we drop the property of countable additivity over the partition, the property $P(P(X|Y)) = P(X)$, need not hold. (This is termed non-conglomerability.)

6. Adjusted belief structures and conditional beliefs

In our treatment of belief structures, we observed that the relationship between prevision and projection is implicit in the definition of prevision. We will now make a similar identification between conditional prevision and the more general projection operator.

Thus, consider a general operator P_B , where B is spanned by the finite collection of elements B_1, B_2, \dots, B_k . By definition $P_B(X)$ is the linear combination $c_1B_1 + \dots + c_kB_k$, where the coefficients are chosen to minimize

$$P(X - (c_1B_1 + \dots + c_kB_k))^2,$$

over all choices of d_1, \dots, d_k . Preferring penalty A to penalty B corresponds to $P(A) < P(B)$. Thus, we may interpret the values c_1, \dots, c_k as the values which you would choose if you were subsequently to suffer the penalty

$$L = (X - (c_1 B_1 + \dots + c_k B_k))^2 .$$

Now consider the definition of conditional prevision, $P(X|H)$. You are required to choose your preferred penalty $H(X-d)^2$, over choices of d , where H is the indicator function of the corresponding event. This looks somewhat different from the choice you have to make in assessing P_H . However observe that when you assess $P(X|H)$, you are also, by coherence, implicitly making a further assessment of $P(X|H_C)$, where $H_C = 1-H$, the complement of H , as

$$P(X) = P(X|H)P(H) + P(X|H_C)(1 - P(H))$$

(so that if you specify $P(X|H)$, $P(X)$, $P(H)$, then this determines the value of $P(X|H_C)$).

Thus, when you consider the penalty $L = H(X-d)^2$, you also implicitly make an assessment for the penalty $L_C = H_C(X-d_C)^2$.

Thus an equivalent formulation of the definition of conditional prevision is that you must specify two values d and d_C and you will incur a penalty

$$L^* = H(X-d)^2 + H_C(X-d_C)^2 .$$

Now as

$$H + H_C = 1, HH_C = 0, H^2 = H, H_C^2 = H_C$$

the above penalty may be identically rewritten as

$$L^* = (X - dH - d_C H_C)^2 .$$

From the discussion at the beginning of this section it is clear that your choice of values $d = P(X|H)$, $d_C = P(X|H_C)$ is equivalent to your choice of the element $dH + d_C H_C$ which is the projection of X into H , the subspace spanned by H, H_C i.e.

$$P_H(X) = P(X|H)H + P(X|H_C)H_C .$$

Notice that because $P(X|H)$ is the coefficient of H in the projection of X into H , we can immediately deduce the usual formula for conditional prevision from the standard formula (5) for the coefficients of the projection operator.

As H and H_C are orthogonal vectors, the coefficient of H in $P_H(X)$ is given by $(X, H)/(H, H)$. As $H^2 = H$, we have that

$$P(X|H) = P(XH)/P(H) ,$$

as required.

Thus, directly from the definition, conditional prevision on H is simply the projection into the subspace H spanned by H and H_C . (Equivalently H is the space spanned by H and X_0 , the unit constant, which explains why your specification of $P(X|H)$ fixes $P(X|H_C)$.)

Of course, if we had first established the relationship $P(X|H) = P(XH)/P(H)$, then we could simply reverse the above argument and deduce the relationship between conditional prevision and projection. Thus, the relationship is not so much a property of our particular choice of definition, but corresponds to any definition which yields the familiar formula for conditional prevision. However, we have preferred to take a formulation in which everything can be immediately deduced simply from a careful statement of the definition itself.

In precisely the same way, if $\Pi = \{E_1, \dots, E_k\}$ is a partition, then letting Π also represent the belief structure spanned by the random quantities E_1, \dots, E_k , we have for any X ,

$$P_{\Pi}(X) = P(X|E_1)E_1 + \dots + P(X|E_k)E_k .$$

Notice that $P_{\Pi}(X)$ is numerically equivalent to the quantity which we termed $P(X|\Pi)$ in section 4. In this sense we move interchangeably between conditional prevision and projection. Notice that this gives a geometric interpretation as to why $P(P_{\Pi}(X)) = P(X)$. That is, as $E_1 + \dots + E_k = X_0$, X_0 is an element of Π , so that $A_0 \subset \Pi$.

Thus, if you determine $P(X)$ by projecting X directly into A_0 , or by first projecting X into Π and then into A_0 , you will obtain the same result in either case. This is simply a special case of the general property $P_B P_D \equiv P_B$ if $B \subset D$.

When we discussed conditional prevision above we observed that it was very disturbing that we appeared to need such a definition, because it appeared to give events a special status that we were anxious to avoid. (For example, in our general discussion of belief

structures we argued against the requirement that we should be forced into a full probabilistic specification for all quantities of interest.)

We have now completed the first step in dispensing with the idea of conditional prevision, namely we have shown that the definition itself does not introduce a new concept into our system, but simply identifies a particular type of projection operator. We may now repeat essentially the same general argument as when we observed that it would be an arbitrary restriction to say that there was something "special" about the prevision of an indicator function i.e. that there was no logical distinction between your consideration of the penalty $(X - \bar{X})^2$ when X was a two valued quantity or when X was a many valued quantity.

In the same way, when you consider your choice of penalty $(X - c_1 B_1 - \dots - c_k B_k)^2$, over choices of c_1, \dots, c_k , there is no logical distinction between your choice when B_1, \dots, B_k happen to be the indicator functions for the events of a partition and your choice when B_1, \dots, B_k are any general random quantities.

Again this is quite separate from the psychological question as to which choices you personally prefer to consider. You may well find it convenient to work with conditional probabilities in certain situations. Notice that if you specify conditional probabilities directly, and deduce various unconditional probabilities from the coherence relations, then this corresponds to a direct specification of the projections into certain subspaces, and construction of the inner product space in such a way as to be consistent with the projections. This illustrates the argument of section 4, namely that the various adjustments of a belief structure offer you a variety of different, but consistent approaches to specifications of your beliefs, and that you should choose the most intuitively meaningful approach for the problem at hand.

It still remains to be considered whether the projection into a subspace spanned by indicator functions has some property (logical, not psychological) which distinguishes it from the more general projection. In particular, it might be thought that the interpretation of a projection into an indicator space is "special", because it corresponds

to the usual Bayesian approach of revising your beliefs by conditioning on observed events (whereas, there is no such obvious correspondence for the general projection).

In later articles we will show that such a distinction is entirely without foundation (at least, in our approach). For now, let us emphasize again that everything that we have said concerning your conditional previsions, and the corresponding projections, relates to probabilistic relationships (expressed now) between various random quantities. There is, as yet, absolutely no implication in anything that we have said to suggest any relationship whatever between your expressed beliefs at different time points. Such relationships will turn out to be the crucial feature of the theory of belief structures. However, because this issue requires very careful consideration, we will defer it completely to later articles when we have laid the necessary groundwork for a proper treatment. Thus, for example, the conditional prevision $P(X|H)$ bears, as yet, no relationship with the value that you may express for $P(X)$ if you learn that H occurs. Do not imagine intrinsic properties of quantities without providing careful justification.

Finally, let us extend the link between belief structures and full Bayesian specifications to cover the general adjusted belief space. Thus, we begin with the space $A = L_2(\Omega, P)$ (i.e. the space of all square integrable functions on Ω under the usual L_2 inner product with respect to the probability measure P). We introduce the new space $B = L_2(S, Q)$, where Q is a probability measure over the probability space S . The projection operator P_g is "conditional expectation", that is for any element $g \in A$, $P_g(g)$ is the element of B defined pointwise for each $s \in S$ by

$$(P_g g)(s) = \int_{\Omega} g(w) dP(w|s) .$$

(The conditioning is with respect to the joint probability distribution on $\Omega \times S$, which essentially creates the space $A + B$.)

In more familiar terms Ω is usually the "parameter space" and S is the "sample space". Typically there is a joint p.d.f. over $\theta \times S$ composed as $f(s, w) = f(s|w)p(w)$, where $p(w)$ is the prior density for the parameter w , and $f(s|w)$ is the likelihood function. The above integral thus reduces to

$$(P_g g)(s) = \frac{\int g(w) f(s|w) p(w) dw}{\int f(s|w) p(w) dw} .$$

7. Adjusted belief construction

As a simple example, to motivate the next stage in our development, let us return to the example discussed in the previous article. Suppose that the teacher decides to restrict the belief space A to quantities for which he is fundamentally interested in specifying his beliefs. In this case, that will be the various numerical measures of the student's performance in the coming year. He places all of the other quantities, such as the previous year's test scores into a second belief structure B (i.e. B contains all of those quantities about which he has no interest in specifying belief except inasmuch as this will suggest or explain or clarify certain of his beliefs over A).

This separation will be an important feature of our further development. Thus, let us term A the primary belief structure and B the support structure. These terms do not reflect fundamental aspects of the random quantities involved (for example, some functions of a particular random quantity may be elements of A , and other functions of the same random quantity may be elements of B). Instead the terms reflect the external objectives for which the specification is being made. We will term $D = A + B$ the total belief structure (i.e. D is the structure which, in principle, delimits the arguments which can be made).

As a simple example, suppose that the teacher has constructed the primary space A spanned by $X_0 = 1$, $X_1 = S$, and the support space B spanned by $\{X_0, Y\}$ (S is the score for the coming test, Y is the score for a comparable earlier test - though the teacher has not yet seen the value of Y). He then constructs the space A/B . Thus, he has assessed values for $P(S)$, $P(S^2)$, $P(Y)$, $P(Y^2)$ and $P(YS)$. He constructs the projection P_B by using the relation $P_B = P_{A_0} + P_{B/A_0}$ and formula (5) to give the familiar "least squares" formulae

$$P_B(S) = P(S)X_0 + \frac{\text{cov}(S, Y)}{\text{var } Y} (Y - P(Y)Y_0) \quad (6)$$

and

$$|P_{A/B}(S)|^2 = |S - P_B(S)|^2 = \text{var } S - \frac{\text{cov}^2(Y, S)}{\text{var } (Y)}.$$

As we have emphasized above, P_B is simply a generalization of P_{A_0} , that is from a numerically fixed prevision (with associated penalty $(S - P(S))^2$) to a numerically random

prevision (with penalty $(S - P_B(S))^2$). If you express a strong preference for $P_B(S)$ over $P_{A_0}(S)$, (as quantified by a large value of $(\|P_{A_0/A}(S)\|^2 - \|P_{A_0/B}(S)\|^2)$, or equivalently a large value of $\frac{\text{cov}^2(S, Y)}{\text{var}(Y)}$), then this information is not qualitatively different from announcing a strong preference for one value of $P(S)$ over another (for example, preferring a penalty $(S-1)^2$ to a penalty $(S-5)^2$). It remains for you to interpret your preference in the context of the problem under consideration. In this case, strong preference for $P_B(S)$ over $P_{A_0}(S)$ might suggest that use of Y to "predict" S but we will consider this in detail in subsequent articles. For now, we view $P_B(S)$ simply as a "random prevision", asserted now.

To illustrate this interpretation, consider the following calculation. Suppose that there is a critical "pass-fail" level s for S , and let $I_S = 1$ if $S > s$ and $I_S = 0$ otherwise. Define $I_Y = 1$ if $Y > y$ (a corresponding critical value), and $I_Y = 0$ otherwise. The obvious quantity to consider is the conditional probability that you will pass test S given that you have passed test Y , i.e. to evaluate $P(S > s | Y > y)$, or equivalently to evaluate $P(I_S I_Y) / P(I_Y)$.

Thus, suppose you assess $P(I_S) = p$, $P(I_Y) = q$, $P(I_S I_Y) = u$ so that $P(S > s | Y > y) = u/q$. The larger the value of u/q compared to p the more "relevant" it may be to observe the event $Y > y$. How can we express this?

From the discussion of section 6 the conditional probability argument can be set in the inner product space. Thus let A be spanned by X_0, I_S , and B by X_0, I_Y . The values p, q, u fully specify the total space $A + B$. Applying formulae (6) we have

$$\begin{aligned} P_B(I_S) &= X_0 P(I_S) + \frac{\text{cov}(I_S, I_Y)}{\text{var}(I_Y)} (I_Y - P(I_Y) X_0) \\ &= p X_0 + \frac{u - pq}{q - p} (I_Y - q X_0) \end{aligned}$$

and

(7)

$$\begin{aligned} \|I_S - P_B(I_S)\|^2 &= \text{var } I_S - \frac{\text{cov}^2(I_S, I_Y)}{\text{var}(I_Y)} \\ &= p - p^2 - \frac{(u-pq)^2}{q-q^2} . \end{aligned}$$

$P_B(I_S)$ takes 2 possible values. If $Y > y$, i.e. $I_Y = 1$, then numerically

$$P_B(I_S) = u/q = P(S>s|Y>y)$$

and if $I_Y = 0$, then numerically

$$P_B(I_S) = P(S>s|Y<y) ,$$

so that, with $\bar{I}_Y = 1 - I_Y$, we have

$$P_B(I_S) = I_Y P(I_S|I_Y) + \bar{I}_Y P(I_S|\bar{I}_Y) .$$

This is of course the formula for projection from section 6. We have derived it again to emphasize that it is precisely the same equation as (6), derived in the same way (i.e. through (7)) and with the same justification (in terms of choosing a "random" prevision) as for (6), but simply applied to two-valued random quantities, rather than many valued random quantities. Just as the norm of the residual vector in (6) plays an important role in determining the value of Y , in assessing S , so does the corresponding norm in (7) relate to the value of I_Y in assessing I_S .

It is up to you whether you want to consider the quantities, summarized in (6) or (7). All that we have observed is that if, for example, you specify the values for $P(I_S)$, $P(I_Y)$ and $P(I_S I_Y)$, then the theory will determine for you the quantities in (7) (and nothing else). It is up to you in any particular problem to decide what quantities you wish to determine. All that theory can provide is an organizing framework in which the implications of your specifications can be clearly displayed. That framework concerns the analysis of belief transformations over A , and will be the subject of our next article.

REFERENCES

- DeFinetti, B. (1974). Theory of Probability, vol. 1, London: Wiley.
- Goldstein, M. (1984). Belief Structures, University of Wisconsin-Madison, Mathematics Research Center Technical Summary Report #2738.

MG/jvs

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #2804	2. GOVT ACCESSION NO. AD-A154804	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Adjusted Belief Structures		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Michael Goldstein		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics & Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P. O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE March 1985
		13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Coherence, conditional prevision, prevision, projection		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) In the study of the foundations of the subjectivist theory of statistics, we find that each aspect of the theory corresponds to a different application of a single manipulation, namely the adjustment of one belief structure by another belief structure. This article describes the technical machinery of this manipulation to a sufficient level of detail to cover all of the applications of the adjusted belief structure in the foundations of the theory. We also discuss the relationship between the adjustment of belief structures and the conditioning of random variables. (Essentially the latter is a simple special case of the		

END

FILMED

7-85

DTIC